CSIS Discussion Paper #2

Recognizing Buildings in Urban Scene of Distant View

Peilin Liu, Katsushi Ikeuchi and Masao Sakauchi
Institute of Industrial Science, University of Tokyo, Japan
7-22-1 Roppongi, Minato-ku, Tokyo 106, Japan
Email:liu@sak.iis.u-tokyo.ac.jp

KEY WORDS: 2D Template Matching, local searching techniques, intensity edgels, pose estimation

ABSTRACT

This paper presents an approach for recognizing urban scene in distant views. Previously, we have developed a method using dynamic programming technique to recognize key buildings appearing as silhouette in a distant view. Based on those locations of buildings in silhouette, this paper develops a method to find locations of other buildings which do not appear as silhouette in the image by using 2D template matching. We then describe some applications of the approach such as incorporating building texture information to 3D building models and estimating the height of abuilding. We also show its effectiveness in experiments on real images.

1 Introduction

With the development of computer science and communication techniques,systems which utilize real world environment information, for example,car-navigation systems, are possible. Collecting, utilizing, and integrating information about the real world environment is important since it can help people to know and grasp what happened in the real world and help them to decide what to do in a new situation. Images are an important source of information about the real world. So it is necessary to construct a system which collects and integrates real world information automatically into a database from images or other sources. The objective of our research is to realize the functions of integrating the information about an urban environment by understanding the urban scene captured by a visual sensor. In urban scenes most of objects are buildings, so understanding urban scenes can be completed by recognizing the buildings and reasoning about the relations between them. This paper proposes an approach for recognizing buildings of urban scenes in distant views.

Recently, database technology, computer graphics techniques, and measurement techniques have been used to build digital city maps. In digital city maps, geometric information on the buildings, such as polygonal shape of the floors and the number of floors, is available. Based on this kind of map, constructing a model of the real world for recognition is possible. In our research, a digital city map is used to build a world model.

We had proposed an approach for recognizing key buildings appearing as silhouette in distant views [1].We described our approach for recognizing buildings from skylines. The skyline consists almost entirely of the contours of the buildings in an urban environment. We select features from the prominent buildings which

might appear as silhouettes in the view for correspondence between the model and images. A city map-database with 3-D descriptions of the rooftops of the buildings is used to build the world model. The rough viewing parameters for location and orientation can be obtained from the sensors. Although these parameters are not precise, the information about which buildings appear as silhouettes can be obtained. Based on this information, a model consisting of buildings' rooftop line segments is constructed for recognition. The feature correspondences between image and the model are established using a dynamic programming technique. The correspondence hypotheses are then verified to ensure that the correspondences are reliable and accurate. Incorrect feature correspondences caused by sensor uncertainty and image clutter are modified, based on a similarity evaluation method. Using this method, the buildings appearing as silhouette in distant views can be identified as shown in figure 1.

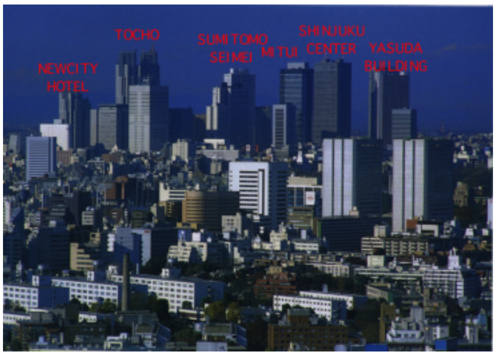


Figure 1: Result of recognition using the approach for recognizing buildings from skylines

In this paper, we further present an approach for finding the true location of the buildings which do not appear as silhouette in the image by using template matching. A city map-database with 3-D descriptions of the rooftops of the buildings is used to build a world model. We first describe an approach for estimating the optimal pose of a building. Based on the accurate pose, we can locate the correct position of building in the image.

Finding the true location of the buildings has some important applications for GIS. We then describe some applications of the approach such as incorporating building texture information to

3D building models and estimating the height of a building. When the template generated by projecting the model to the image is matched with the building in the image, the texture information of the building can be obtained and incorporate to the digital map. At present, the number of floors of a building is the only information on the height of a building which can be obtained from a digital city map. One may want to know the exact height of the buildings in order to construct exact 3D building models to carry out a recognition task. In this paper, we also present an approach for estimating the height of a building accurately by localizing an reference building whose height is assumed to be known. We also give some experimental results and compare the experimental results with other height estimation methods and see the effectiveness of the approach which measures a building's height by 2DTM.

2 Locating Buildings in Urban Scenes in Distant Views

Based on the result of the corresponding building features between the image and the model, we further present an approach for locating the buildings which are not visible in silhouette in the image by using template matching, so that the contents of the scene can be understood more completely. The building locations can be estimated by projecting their models onto the image if the projecting transformation is known. The transformation from the model to the image can be computed from the corresponding model and image points obtained by the approach we described in previous section. The projected building models will match the buildings in the image if the location estimates is accurate enough. Unfortunately, because of slight inaccuracies of the location estimates, the probability that the projected building nmodels will not match with the buildings in the image is high. The solution adopted here is to use this location estimates as starting points for a local search for the true locations of the buildings. Local searching techniques[3][4][5] have been widely used in computing accurate object pose, and have been regarded as complimentary methods for some recognition algorithms. 2-dimension template matching (2DTM) presented by [2] is one such algorithm. In this approach, intensity edgels(edge element along an edge chain and its tangent direction) is used as the primitive feature for matching the model to the image, and the pose is evaluated by measuring the distance between points on the projected model and points in the image. In this section, we describe the approach for locating buildings in the image by using 2DTM. First, we use the 3D building models to predict the visible edgels of the model given the pose. Then we compute the correspondences between the visible model edgels and image edgels based on nearest-neighbor search in 2D. Once the correspondences are established, the accurate pose of the building can be obtained. Getting the accurate pose of buildings has some important applications for GIS, such as taking the texture of buildings and pasting them onto 3D models, which will be detailed in the next section.

2.1. 3D Building Models for 2DTM

In 2DTM, the object model is represented as a collection of 3D edgel generators which are points on the object surface which often create edges visible in intensity images. Using the edgel generators the 2DTM system can predict the appearance of object edgels in an image and then match them with intensity edgels in the input image.

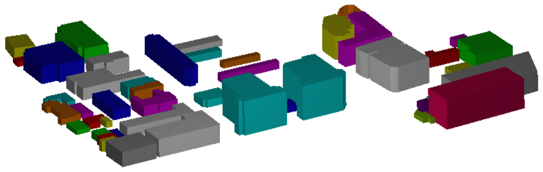


Figure 2: An image of a digital map

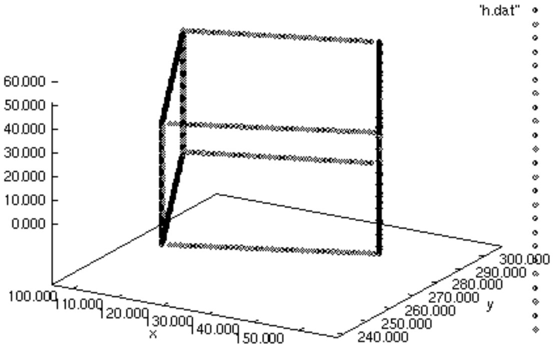


Figure 3:Model of Honda Building

The 2D geometric shape of the floors and the number of floors of buildings is available from digital map, like the one as shown in Figure 2. The 3D building models for 2DTM can be constructed based on this geometric information. The height of buildings can be inferred by defining the height of each floor. Since the convex contour of buildings are prominent in the distant views, we construct edgel generators just using the convex contour edges of the building. The convex contour edges are constructed by 3D points in world coordinates along with the tangent direction for each point. Figure 3 shows one example of a building model for 2DTM.

2.2 Computing the Initial Pose of a Building

2DTM is an approach for computing the precise pose of a 3D object in a 2D intensity image given a rough estimate of the object pose in the image. The rough estimate of the object pose can be computed based on the correspondences obtained in the previous section. We have three coordinate systems to deal with: image, camera, and object. Under the perspective projection model of the imaging process, a transformation from a solid model to an image can be computed from six pairs of model and image points, shown in equation(1). C is the camera transformation. Known one pair of the correspondence points of image X_i and model X_o, two parameters of the camera transform can be computed. If six pairs of the correspondences not in the same surface are given, the camera transformation can be obtained.

$$X_i = X_o \cdot C \qquad (1)$$

The equations for solving this problem are relatively unstable, and the most successful methods use more than six points and an error minimization procedure such as least squares see equation(2).

$$\mathrm{C} = (\mathrm{X}_o^t\ \mathrm{X}_o)^{-1}\ \mathrm{X}_o^t\ \mathrm{X}_i \qquad (2)$$

We compute the camera transformation to compute the transformation by equation(2) using the endpoints of the line segments from the corresponding building models and the buildings in the image. The pose of an object can be defined by three rotation parameters and three translation parameters. With the transformation, the initial pose of the building with respect to the camera can be calculated directly.

2.3. Locating a Building using 2DTM

The visibility of points of the building can be computed from the initial pose. With these visible points on the projected model, we can compute correspondences between the model edgels and edgels in the image. The canny edge operator and an edgel linker is applied to the intensity image, and the resulting edgel chains are smoothed to remove alias and noise effects. The result is a set of smooth 2D edgel chains in image coordinates. To find the nearest neighbors in 2D image space, we project the visible 3D model edgels into 2D image coordinates and then search for their nearest neighbors using the k-d tree method. We have a rough pose estimate and many correspondences between model points and image points. Most of these correspondences will be incorrect initially, because of using the weak projective imaging model or because of the pose estimate. The location algorithm used here is iterative process which refines the pose by optimizing an objective function defined over the image data, model data and the building's pose. The building model and the building in the image are matched well when the optimal pose is obtained. We can find the optimal pose by minimizing this objective function:equation: 3.

$$E(q) = \sum_{i} \rho(Z_i(q)) \qquad (3)$$

Here Z_i is the error of the ith model and image correspondence given by equation(4)

$$Z_i(q) = \min_{\vec{a} \in D} \| \vec{x}_i - \vec{a} \| \qquad (4)$$

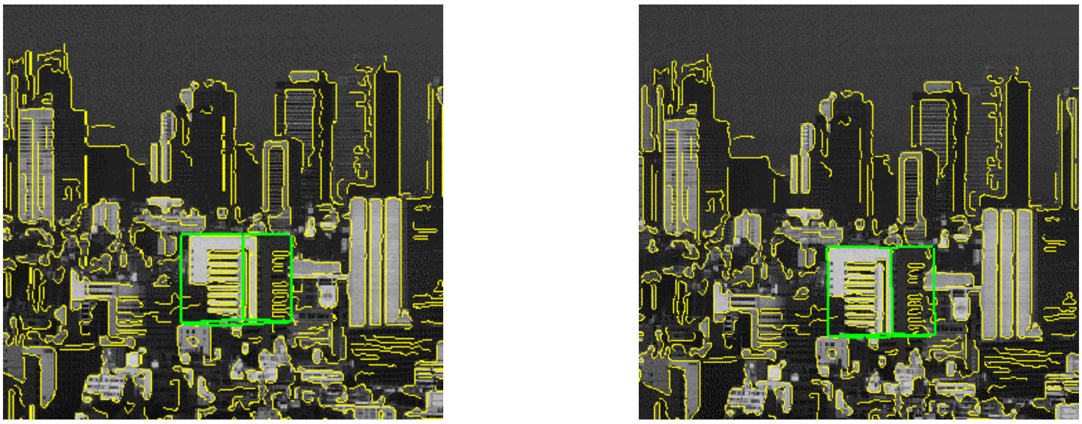
where D is the set of 3D data points in the image, $\vec{x}_i$ is the 2D image coordinate of the ith model edgel point transformed using the camera parameters q. $\vec{a}$ is nearest image point to the projected model point, and is found] by using a k-d tree nearest-neighbor search. We assume that the values of z are independent samples of the distribution P(z)$\propto e^{-\rho(z)}$. To minimize the objective function, we take steps in q along the gradient direction, according to equation(5).

$$\Delta q \propto -\frac{\partial E}{\partial q} = \sum_{i=1}^{n} \Psi(z_i) \frac{\partial z_i}{\partial q} \qquad (5)$$

Here $\Psi(z_i) = \frac{d\rho(z)}{dz}$. The correspondences are recomputed at each function evaluation. This process ends when a minimum-size step in the gradient direction no longer improves the estimate, at which point the accurate pose can be regarded to be obtained. Based on the accurate pose, we can locate the correct position of building in the image.

2.4. Experimental Results

We then used 2DTM to locate the Honda building in the image as shown in Figure 4, using the model generated based on map information shown in Figure 3. As figure 4 shows, the projected Honda Building model initially did not match the building in the image well. With 2DTM, we can find the exact position of the building. Figure 4 shows the matching result.



a)before matching b)after matching

Figure 4:Locating a building by 2DTM

3. Applications of Locating buildings using 2DTM

Finding the true location of buildings has some important applications in GIS. For example, when the template generated by projecting a model onto an image is accurately matched with the building in the image, the texture information of the building can be obtained and incorporated into the 3D building model. Thus we can automatically add texture information to the building model database, to support recognition tasks in which texture information is needed. In addition, one may want to obtain 3D building models to carry out recognition of buildings. At present,the number of floors of a building is the only information on the height of a building which can be obtained from a digital map. We can compute the height of a building accurately by localizing other buildings whose heights are assumed to be known. We will describe how to realize these applications in the following sections.3.1. Incorporating Building Texture Information To 3D Building Models

Using 2DTM, when the buildings in an image are matched with projected building models, the visible surface of the buildings can be extracted using the coordinates of the projected models. Figure 5 shows the visible surface of Honda building extracted after accurately localizing the position of the building.

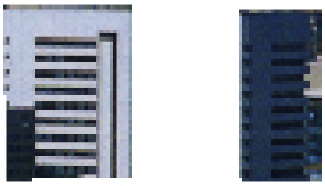


Figure 5: The visible surface of Honda building

With the surface extracted from the image, this texture can be incorporated to the building models, as shown in Figure 6. Incorporating building texture into the 3D building model can help people make 3D GIS systems in which information on buildings is not only geometric but also includes texture.

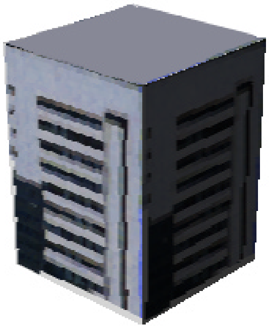


Figure 6: An example of incorporating building texture information into 3D building models

3.2 Estimating the Height of a Building

At present, the number of floors of a building is the only information on the height of a building which can be obtain from a digital city map. Although the height of the building can be estimated from the number of floors, for example by multiplying by an assumed height of each floor, the error may be large.

One may want to know the exact height of the buildings in order to construct exact 3D building models to carry out a recognition task. Here this paper presents an approach for estimating the height of a building accurately using 2DTM to localize an reference building whose height is assumed to be known. We finally give some experimental results and compare the experimental results with other height estimation methods and see the effectiveness of the approach.

3.2.1 An approach of estimating the Height of an object

We have described an approach for estimating the optimal pose of an building using template matching. The optimal pose can be established by minimizing an objective function, see equation(3). Thus, the output of the function E is a measurement of how well the projected model and the object in the image are matched. When the optimal pose of reference building is found, the pose of the building to be measured must be optimal. But since only the number of floors of this building is known, its model is not accurate. Although the height of the building can be estimated from the number of floors, for example by multiplying by an assumed height of each floor, the error may be large. We vary the height of the building to be measured, and when the height reaches the correct value, the output E of objective function of the measured building will be minimal.

3.3.Experiments and Investigation

We give an example, shown in Figure 7, to illustrate the effectiveness of our approach. In Figure 7, we will estimate the height of the building on the right named A, assuming the height of the building on the left named B is known. We first project the models of A and B onto the images as shown in Figure 8. Due to the error of camera calibration and the error of location measurement of the buildings in real world, the projected models do not match the ones in the image.

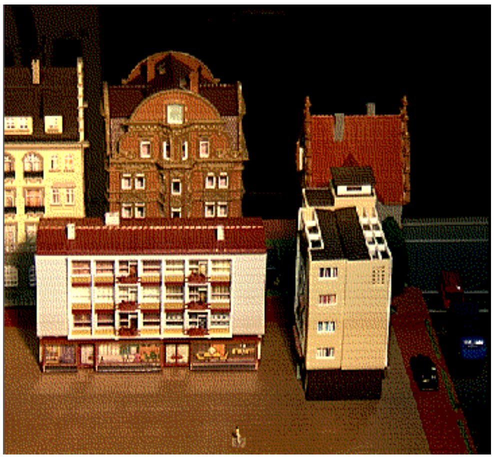


Figure 7: Estimating the height of the building A on the right when the height of the building B on the left is known.

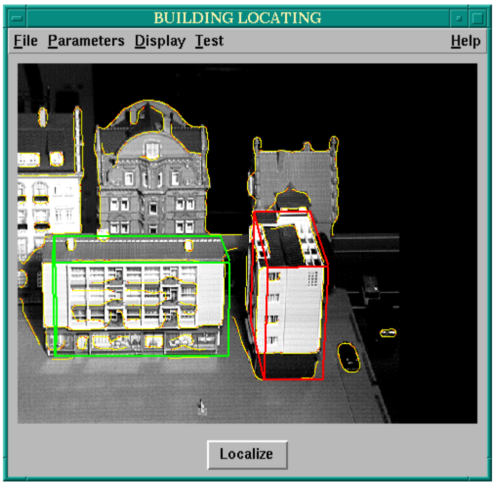


Figure 8: The initial pose of the buildings, assuming that the exact 3D model of B and the number of floors of A are known

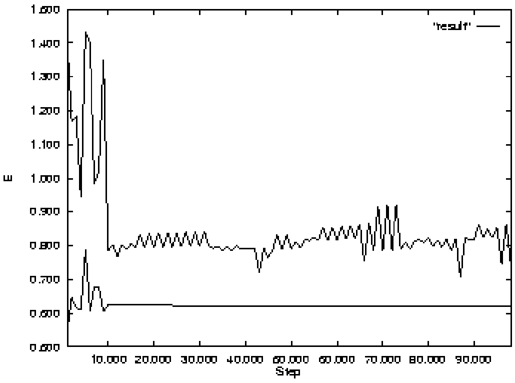


Figure 8: The relationship between the pose of the reference building and the pose of the one which will be measured

We investigate the relationship between the pose of the reference building and the pose of the one which will be measured, as shown in Figure 9. Figure 9 shows when the pose of reference building is optimal, the E of the other one is small.

Height	E	Height	E	Height	E
75	2.1907	88	1.9663	93	2.1086
80	2.0323	89	1.9375	94	2.1877
83	2.0027	90	2.0554	95	2.1371
85	1.9933	91	2.0846	96	2.1714
86	1.9777	92	2.0881	100	2.2965

Table 1: Variation of Height and E of Building A

For each height, we localize 100 times, record the output of objective function of the object, and compute the average of them. In Table 1, E is the average value of the output of the objective over

100 trials. As shown in Table 1, E is the smallest when the height of A is 89 mm. Figure 10 shows the projected model A matched with the object A after localization.

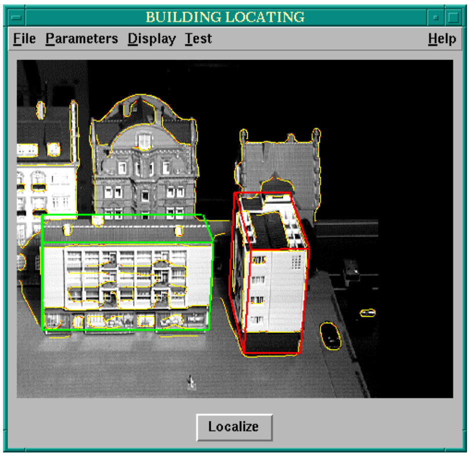


Figure 10:Matching result of building A

3.4 Precision

We compare the experimental results with other height estimation methods and see the effectiveness of the approach which measures a building's height by 2DTM. One method is to estimate the height of buildings from the number of floors. In this experiment, the height of buildings can also be obtained from a CAD model. The height value obtained from CAD model can be regarded as accurate. Table 2 shows the comparison of the results.

Building	Standard	2DTM	Error	Num. X 20	Error	Num.x15	Error
A	93	89	4	5x20	7	5x15	18
B	83	80	3	4x20	3	4x15	13

Table 2: Comparison of methods for estimating the height of the buildings

From Table 2, we can see that the error in the height measured using 2DTM is smaller than the error in estimating the height from the number of floors.

4.Conclusion

In this paper, we have described our approach for understanding urban scenes in distant views. In this approach, the buildings appearing as silhouette in a distant view are first recognized using dynamic programming technique. Based on the result of the correspondence of building features between the image and the model, the camera parameters are computed. The visibility of points of the building to be recognized can be computed from the camera parameters. With these visible points on the projected model, correspondences between the model edgels and edgels in the image are computed. Once the correspondences are established, the accurate pose of the building can be obtained. Getting the accurate pose of buildings has some important applications for GIS, for example, when the template generated by projecting the model to the image is matched with the building in the image, the texture information of the building can be obtained and attached to the digital map and the 3D information on buildings can be estimated. Moreover if a building which should be in the scene is not detected, it can be assumed that construction work or a disaster has happened there.

Using this system, you will be able to know what buildings you see and get some information about them automatically. It can also be used in monitoring urban areas to tell people what happened in there, for example the situation of a place where an earthquake or other natural disaster happened, and to help people to decide what to do with the new situation. Such a system can make the life of people convenient.

Reference

1.Peilin Liu, Wei WU, Katsushi Ikeuchi, Masao Sakauchi, ``Recognition of Urban Scene Using Silhouette of Buildings and City Map Database", ACCV'98(1998.1)

2. Mark D. Wheeler and Katsushi Ikeuchi ``Sensor Modeling, Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition" IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.17, No.3,pp252-265,1995

3. C.Fennema, A. Hanson, E.Riseman, J.R. Beveridge, and R. Kumar, "Model-Directed Mobile Robot Navigation," IEEE Trans. on Syst., Man, and Cybe., Vol.20, no.6, pp.1352-1369, Nov./Dec. 1990.

4. J.R. Beveridge and E. M. Riseman,"How Easy is Matching 2D Line Models Using Local Search," IEEE Trans. on Patt. Anal. Machine Intell., Vol.19, no.6, pp. 564-579, 1997.

5. R.Kumar and A.R.Hanson,"Robust Methods for Estimating Pose And A Sensitivity Analysis," Proc. CVGIP: Image Understanding, Vol.11,1994.